И.И. Гладкий, Л.П. Махнист, В.С. Рубанов, Т.Ю. Юхимук
Беларусь, Брест, БрГТУ

О ВЫБОРЕ ШАГА ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ПРЯМОГО РАСПРОСТРАНЕНИЯ

Задача обучения нейронной сети состоит в нахождении весовых коэффициентов w_{ij} и порогов T_j нейронной сети, которые минимизируют функцию ошибки сети

$$E(w_{11},\ w_{21},\ \dots,\ w_{m1}, T_1,\ \dots,\ w_{1n},\ w_{2n},\ \dots,\ w_{mn}, T_n) = \frac{1}{2}\sum_{j=1}^{n}\left(y_j - t_j\right)^2$$

где $y_j = F\left(S_j\right)$ – значение функции активации j-ого выходного нейрона сети, $S_j = \sum_{i=1}^{m} w_{ij}x_i - T_j$, x_i – выходное значение i-ого нейрона предыдущего слоя, t_j – ожидаемый выход j-ого выходного нейрона ($i = \overline{1,\ m}$, $j = \overline{1,\ n}$).

Введем обозначения:

$\overline{W} = (w_{11},\ w_{21},\ \dots,\ w_{m1}, T_1,\ \dots,\ w_{1n},\ w_{2n},\ \dots,\ w_{mn}, T_n)^T$ – вектор-столбец весовых коэффициентов w_{ij} и порогов T_j нейронной сети, а $\overline{W}_j = \left(w_{1j},\ w_{2j},\ \dots,\ w_{mj}, T_j\right)^T$ – вектор-столбец весовых коэффициентов w_{ij} и порога T_j, связанных с j-ым выходным нейроном сети, $E\left(\overline{W}\right) = \frac{1}{2}\sum_{j=1}^{n}\left(y_j - t_j\right)^2$ – функция ошибки сети, $E\left(\overline{W}_j\right) = \frac{1}{2}\left(y_j - t_j\right)^2$ – функция ошибки j-ого выходного нейрона сети.

Обучение нейронной сети с использованием метода наискорейшего спуска состоит в изменении весовых коэффициентов w_{ij} и порогов T_j нейронной сети на каждом $(t+1)$ шаге обучения ($t = 1,\ 2,\ \dots$) в соответствии с формулами:

$\overline{W}_j(t+1) = \overline{W}_j(t) - \alpha_j(t)\nabla E\left(\overline{W}_j(t)\right)$, если шаг обучения $\alpha_j(t)$ выбирается только для минимизации функции ошибки сети $E\left(\overline{W}_j\right)$, где $\nabla E\left(\overline{W}_j\right)$ – градиент функции $E\left(\overline{W}_j\right)$,

$\overline{W}_j(t+1) = \overline{W}_j(t) - \alpha(t)\nabla E\left(\overline{W}_j(t)\right)$, если шаг обучения $\alpha(t)$ выбирается для минимизации функции ошибки сети $E\left(\overline{W}\right)$.

При некоторых условиях на функции $E\left(\overline{W}_j\right)$, (например, в [1]) предлагается использовать соотношения:

$$\alpha_j(t) = \frac{\left\|\nabla E\left(\overline{W}_j(t)\right)\right\|^2}{\left(\nabla^2 E\left(\overline{W}_j(t)\right)\cdot\nabla E\left(\overline{W}_j(t)\right),\ \nabla E\left(\overline{W}_j(t)\right)\right)}, \quad (1)$$

$$\alpha(t) = \frac{\sum_{j=1}^{n} \left\| \nabla E\left(\overline{W}_j(t)\right) \right\|^2}{\sum_{j=1}^{n} \left(\nabla^2 E\left(\overline{W}_j(t)\right) \cdot \nabla E\left(\overline{W}_j(t)\right),\ \nabla E\left(\overline{W}_j(t)\right) \right)}, \tag{2}$$

где $\left\| \nabla E\left(\overline{W}_j(t)\right) \right\|$ – длина вектора градиента $\nabla E\left(\overline{W}_j(t)\right)$, связанная со скалярным произведением $\left(\nabla E\left(\overline{W}_j(t)\right),\ \nabla E\left(\overline{W}_j(t)\right) \right)$, а $\nabla^2 E\left(\overline{W}_j(t)\right)$ – матрица Гессе функции $E\left(\overline{W}_j(t)\right)$. Соотношения (2) получены и использовались, например, в [2–4].

Величины $\alpha_1(t)$, $\alpha_2(t)$,…, $\alpha_n(t)$ и $\alpha(t)$, связаны соотношением $\alpha(t) = \frac{1}{\sum_{j=1}^{n} \frac{q_j(t)}{\alpha_j(t)}}$, где $q_j(t) = \frac{\left\| \nabla E\left(\overline{W}_j(t)\right) \right\|^2}{\left\| \nabla E\left(\overline{W}(t)\right) \right\|^2}$, т. е. $\alpha(t)$ – взвешенное среднее гармоническое величин $\alpha_j(t)$, с весами $q_j(t)$.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Maxnist, L. Convergence Analysis of Neural Networks Training Based on steepest Descent Method / L. Maxnist, A. Doudkin, V. Golovko // Pattern Recognition and Information Processing (PRIP’2007) : Proceedings of the Ninth International Conference, Minsk, Republic of Belarus, 22–24 May 2007 : in 2 vol. – Minsk, 2007. – Vol. 1. – P. 285–289.

2. Maniakov, N. Traing algorithm for forecasting multilayer neural network / N. Maniakov, L. Makhnist, V. Rubanov // Pattern Recognition and Information Processing : Proceedings of The Seventh International Conferences (PRIP’2003), Minsk, Republic of Belarus, 21–23 May 2003 : in 2 vol. – Minsk, 2003. – Vol. 1. – P. 26–30.

3. Golovko, M. Multilayer neural networks training methodic / M. Golovko, L. Makhnist, N. Maniakov // Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS’2003) : Proceedings, Lviv, Ukraine, 8–10 Sept. 2003. – Lviv, 2003. – P. 185–190.

4. Makhnist, L. Some Methods of Adaptive Multilayer Neural Network Training / L. Makhnist, N. Maniakov // International Journal of Computing. – 2004. – Vol. 3. – P. 99–106.